(12) UK Patent Application (19) GB (11) 2 285 359 (13) A

(43) Date of A Publication 05.07.1995

(21) Application No 9326582.5

(22) Date of Filing 31.12.1993

(71) Applicant(s)
Philips Electronics UK Limited

(Incorporated in the United Kingdom)

420-430 London Road, CROYDON, CR9 3QR, United Kingdom

(72) Inventor(s)
Vassilis Seferidis

(74) Agent and/or Address for Service
R J Boxall
Philips Electronics, Patents and Trade Marks Department, Cross Oak Lane, REDHILL, Surrey, RH1 5HA, United Kingdom

(51) INT CL^6
H04N 7/32 13/00 // H04N 5/14

(52) UK CL (Edition N)
H4F FD12X FD15 FD3 FD31X FRX

(56) Documents Cited
GB 2198310 A EP 0353644 A2 EP 0254643 A2
EP 0005918 A1

(58) Field of Search
UK CL (Edition N) H4F FDD FGM FRG FRP FRW FRX
INT CL^5 G06F 15/70, H04N 5/14 7/137 13/00 15/00
Online databases: WPI

(54) Disparity coding images for bandwidth reduction

(57) First and second images (such as the left and right images of a stereo pair) are initially divided (30, figure 2) into corresponding blocks of a first size and, for each block of the first image, a search is made (32) for a block at corresponding and nearby positions in the second image. If the search is unsuccessful, the first image block in question, and the blocks searched in the second image, are subdivided and the search repeated for each sub-block. When a reasonable match is made, the required shift is stored and a series of distortions (34) - see figure 4 - are applied to the selected second image block to identify (36) which distortion, pattern, if any, improves the match. The resulting data, from which the first image may be recreated comprises, for each first image block or sub-block, identification of the selected second image block or sub-block in the second image and the shift and distortion applied.

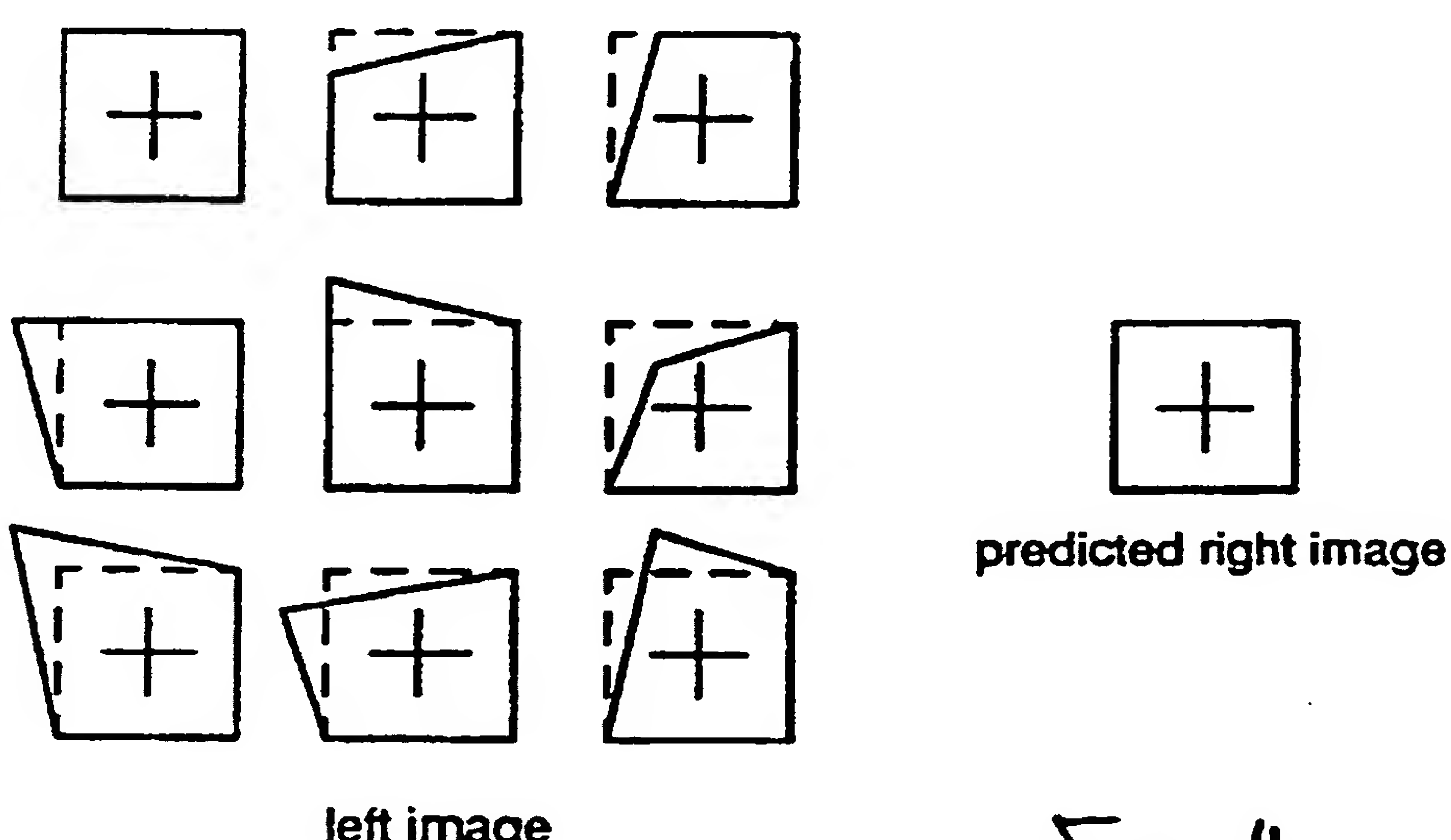

Fig. 4

GB 2 285 359 A

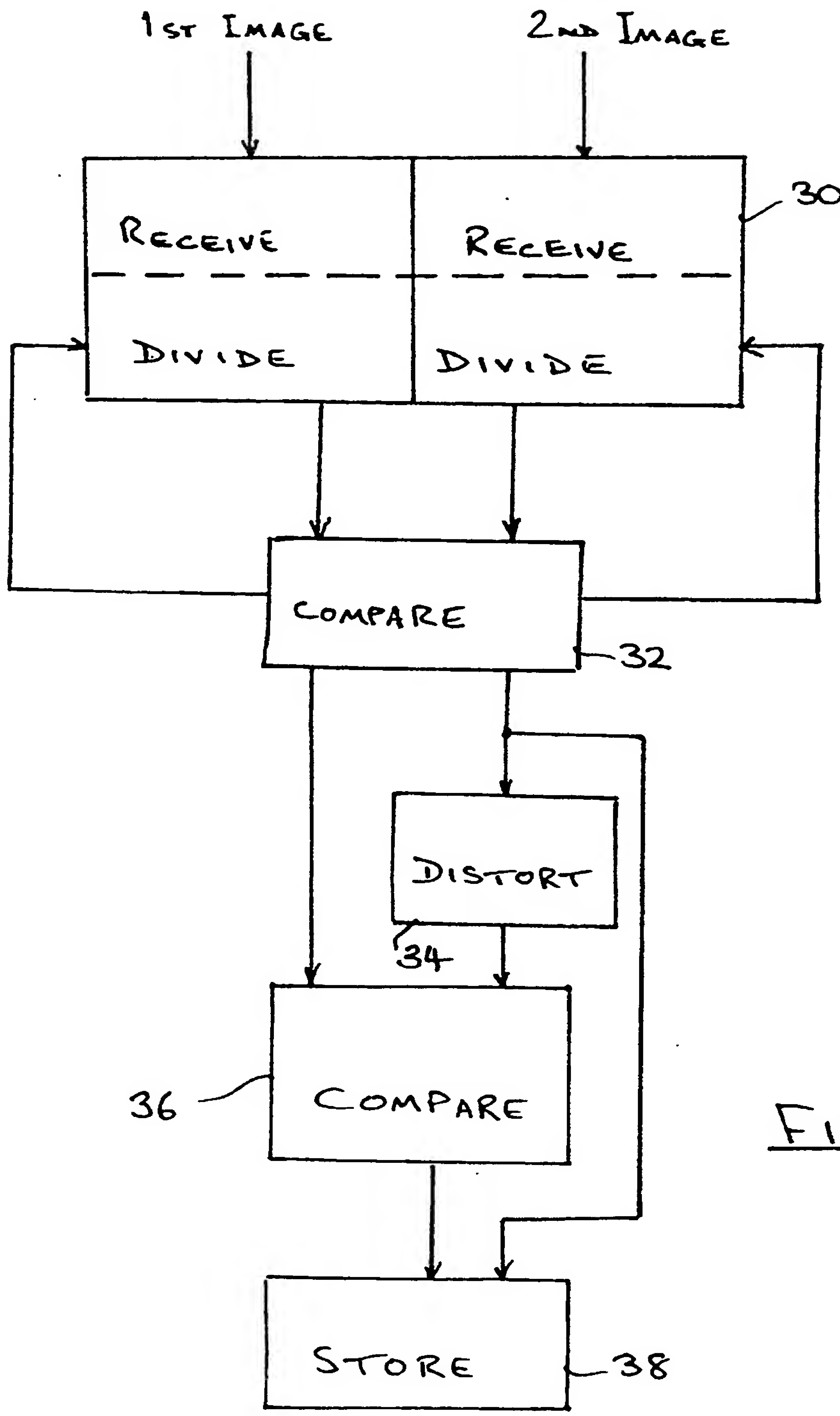

Fig. 2

PHB 33890

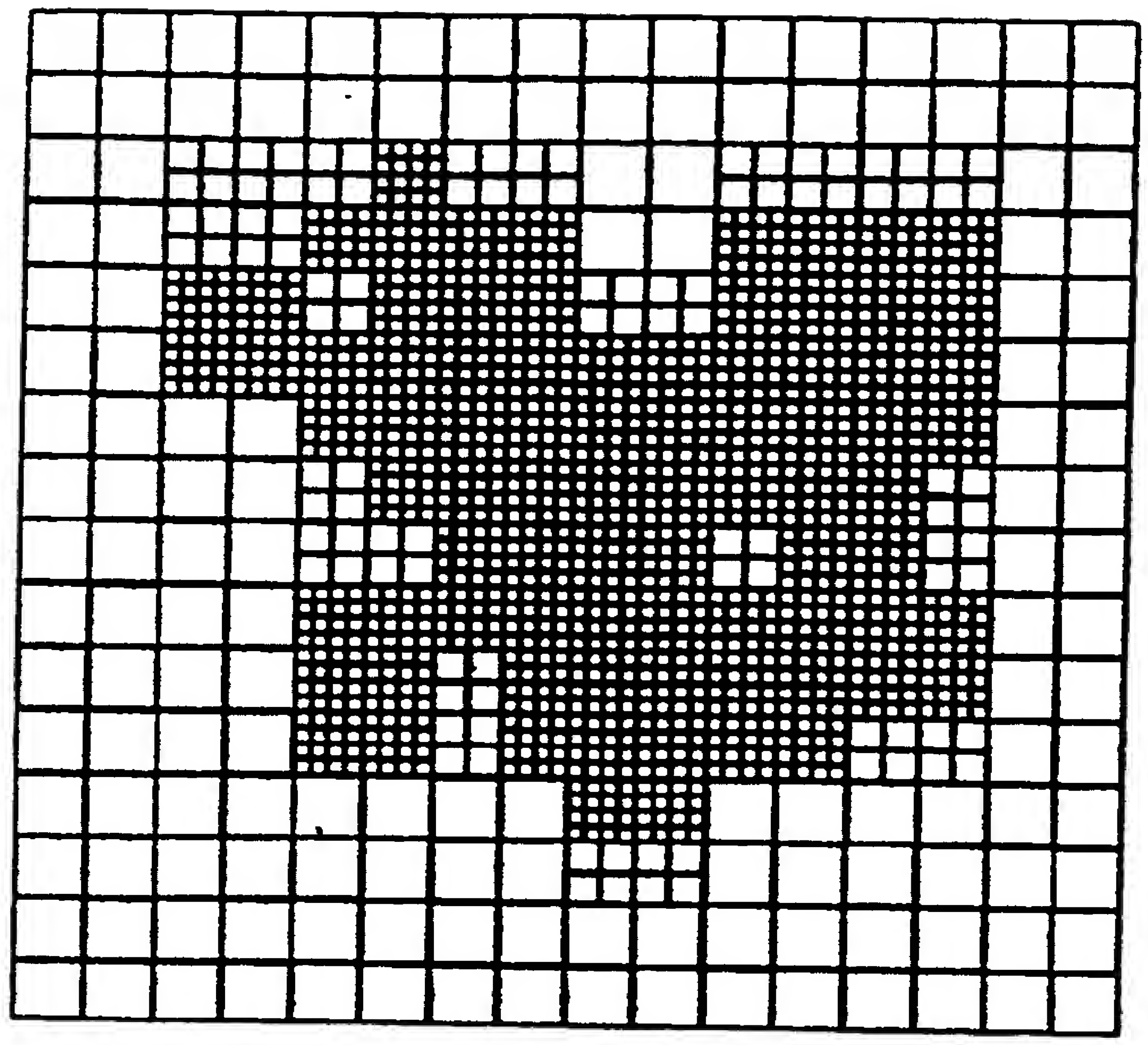

FIG. 6

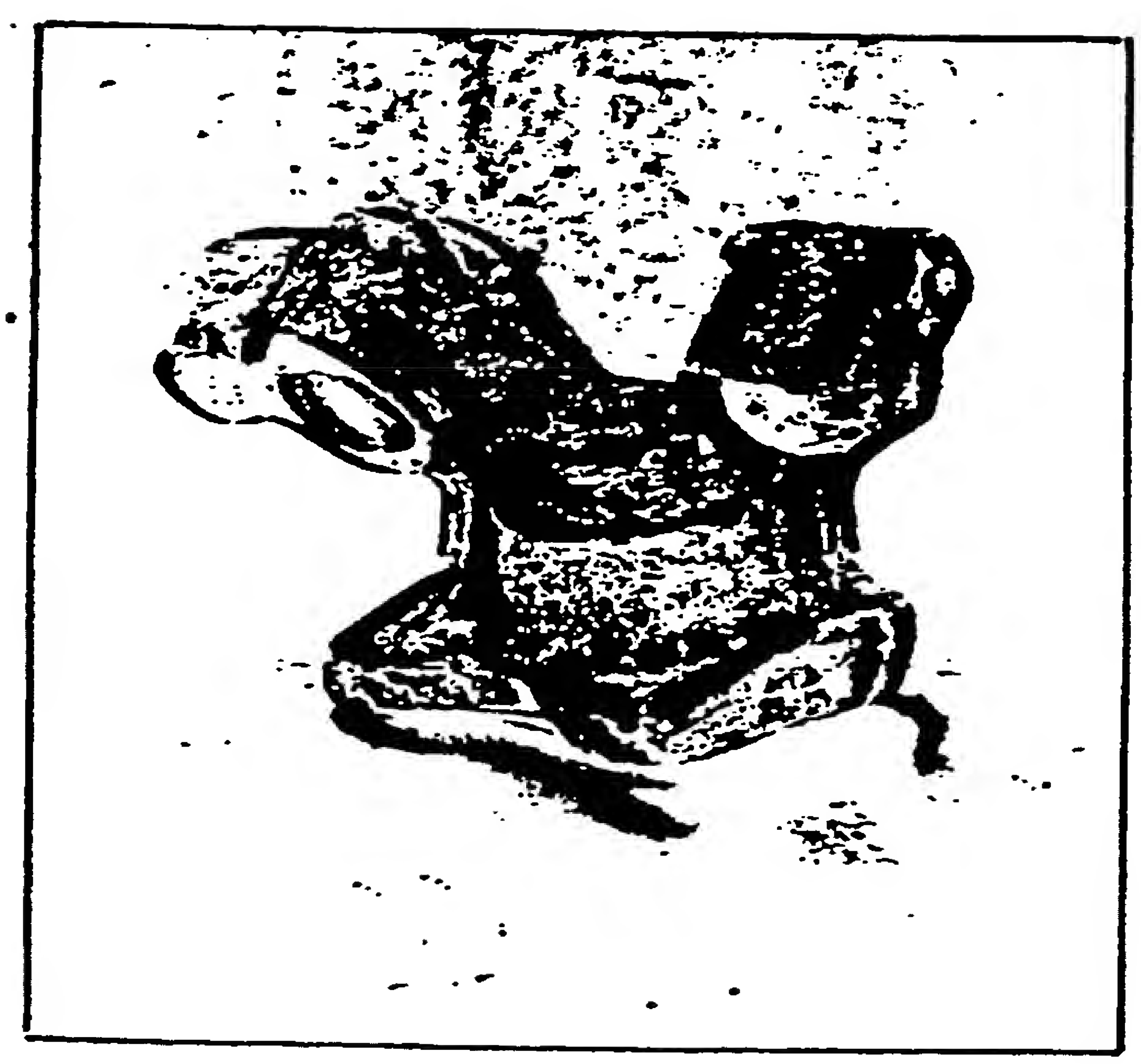

FIG. 7

PHB 33890

areas (pixel blocks) satisfy a selection criterion. Disparity estimation on the other hand must always determine the correct disparity vector, since one stereo image will be reconstructed from the other, at the receiver, and used for stereo perception. Hence, incorporation of convention block based vector estimation techniques have been found unsatisfactory, mainly due to a blockiness effect which degrades the picture quality and forbids the preservation of depth perception. The problems of incorporation are discussed in greater detail in "Improved Disparity Estimation in Stereoscopic Television" by V. Seferidis and D. Papadimitriou, Electronics Letters, Vol.29, No.9, April 1993, pp. 782-783.

Image formation may be considered as a mapping process in which the three-dimensional (3D) scene space is projected onto the two-dimensional (2D) image plane. Due to the many-to-one nature of the mapping, the 3D depth information is lost after projection. The depth ambiguities in the resulting 2D image not only give rise to many problems in scene analysis and image understanding applications, they also eliminate the cues for the determination of the spatial relationships between points and surfaces in a scene. Stereo vision provides a direct way of inferring the missing depth information by using two images (a stereo pair) destined for the left eye and right eye respectively.

The stereo images are generated by recording two slightly different view angles of the same scene. Typically, the stereo camera arrangement consists of two identical cameras which are placed close to each other on the same baseline. When each image of a stereo pair is viewed by its respective eye, the stereo image is perceived in 3D. The differences between the two images are very important since they embody

the estimated display field.

For stereo image coding this problem becomes even more important because it affects the good reconstruction of one stereo image from the other. A possible solution is to interpolate the missing disparity vectors assuming a monotonic variation of disparity values between the existing samples. The interpolation however increases the computational load without securing a better overall performance. Therefore it has been argued that the requirements of a stereo coding system favours the adoption of area-based methods rather that feature-based ones. Moreover an area-based stereo coding scheme simplifies the design of coders compatible with existing video coding standards such as H.261, MPEG I, MPEG II. This is very important if disparity and motion estimation algorithms are to be combined in order to exploit temporal as well as spatial similarities of the two sequences.

An additional drawback of traditional matching methods is their poor performance in handling occluded areas, that is to say those parts of the scene which are present in one image only. The problem is worse in stereo compared to interframe occlusion (sometimes referred to as uncovered background) because the twin camera arrangement introduces its own geometric deformations. The deformations are impossible to be compensated with traditional block matching algorithms because they inherently estimate only translations. A successful disparity estimation scheme must be able to cope with non-linear deformations and occluded objects in order to reconstruct accurately the scene depth. We have recognised that one method which inherently has these properties is generalized block matching, which provides a method for disparity encoding as set forth in the

criteria are satisfied, and the steps of comparing, applying distortions and storing are then performed for each block or sub-block meeting the criteria.

By applying the generalized block matching scheme to blocks of differing sizes, the present invention compensates more accurately those parts of a scene having large disparity values due to the allocation of smaller blocks to those areas.

The division of blocks into sub-blocks, and the subsequent division of sub-blocks into further sub-blocks, may suitably comprise dividing into four equal portions of the same shape as the parent. To prevent excessive computation to provide minimal effect, a minimum sub-block size may be specified such that the matching criteria are assumed to be satisfied when this block size is reached, regardless of whether or not a further subdivision would otherwise be indicated.

Where each block or sub-block is made up of a number of pixels, the distortions applied by generalized block matching may comprise sequentially moving each corner of a block or sub-block to a number of different positions about its original position, such as a pattern of nine positions each spaced from the original position by n pixels in the horizontal, vertical or diagonal direction, where n is an integer.

Following identification of the applied distortion providing the best match, the mean absolute difference between the first image block and the undistorted second image block and between the first image block and the distorted second image block may be compared and, where the ratio between the two mean absolute differences is below a

and operable to receive and store, for each block or sub-block of the first image, information identifying the block of the second image meeting the predetermined matching criteria, and the applied distortion best satisfying the second predetermined matching criteria.

Further features and advantages of the present invention will be apparent from reading of the claims and the following description of preferred embodiments of the present invention, now described in the context of stereo image coding by way of example only and with reference to the accompanying drawings in which:

Figure 1 is a block diagram of a stereoscopic television coding scheme;

Figure 2 is a block schematic representation of coding apparatus embodying the present invention;

Figure 3 illustrates the principle of generalized block matching;

Figure 4 illustrates the matching of quadrilaterals using a 9-position search algorithm;

Figure 5 shows the principle of quad-tree segmentation;and

Figures 6 and 7 respectively represent the absolute difference and quad-tree segmentation for a stereo pair of a test image.

In Fig.1 there is shown a stereoscopic television coding scheme which incorporates motion/disparity compensation. Left and right image sequences 10,12 are applied initially to respective motion estimators 14,16 which eliminate the redundancy between successive frames of the same sequence. The resulting images are then passed to a

passes details of the selected second image block or sub-block and the applied distortion to a buffer 38 for subsequent storage or transmission.

The technique of distorting and comparing is known as generalized block matching and is a block matching technique which approximates the deformations of real objects by deforming the corresponding blocks in the picture. As with other block matching techniques, the image is divided into non-overlapping square blocks and a multi-dimensional vector is assigned to each one. For a stereo image pair, the vector consists of the mapping parameters which satisfy the following criteria:

$$x_i^r = f_1\,(x_i^l, y_i^l)$$

$$y_i^r = f_2\,(x_i^l, y_i^l)$$

$$\Sigma_{i=1}^{N^2}\,(g_i^r - g_i^l)^2 = \min$$

where g_i^r and g_i^l represent the grey values of two blocks of NxN pixels each, of the right and left image respectively and x_i^r, y_i^r and x_i^l, y_i^l are the corresponding x and y coordinates, where $i = 1,2,\ldots.N^2$. As will be appreciated, the above summation represents the mean-square-error criterion which guides the search for the optimal position, although other distortion or correlation measures may be used instead. The mapping functions f_1 and f_2 relate the coordinates of the corresponding images in the two stereo images. It is not necessary for f_1 and f_2 to be linear or monotonous: they may represent one-to-many mappings in order to compensate the non-linear deformations introduced by the stereo arrangement. For backwards compatibility with existing block matching techniques however, it is desirable for both functions to

calculation of these parameters are described in the above-mentioned Optical Engineering article of V. Seferidis and M. Ghanbari. The idea of fast search algorithms is to selectively check only a small number of possible search positions assuming that the distortion measure monotonically decreases towards the best match position. Hopefully, by checking only some representatives of the whole set of possible combinations, the same accuracy can be achieved but with only a fraction of the operations. An example of such a search is shown in Fig.4 for a block of 16x16 pixels: for the sake of simplicity, only variations of the top left-hand corner of the matching block in the left-eye image are shown. Each quadrilateral is formed by displacing the top left corner by +/- 4 pixels horizontally, vertically or diagonally. For each displacement, all the remaining three corners are similarly displaced and the quadrilateral which minimises the mean-square-error is chosen as the best match. It is easy to verify that the total number of quadrilaterals from the left-eye image that are matched with a square block on the right-eye image is $9^4 = 6561$.

As with all block matching techniques, the performance of generalized block matching increases with a reduction of the block size, for example using 8x8 pixel blocks rather that 16x16. However the small block size increases the overhead information (mapping parameters) that must be transmitted or stored. A possible solution to alleviate this problem is to segment the areas that exhibit uniform disparities and to estimate the mapping parameters for each one separately. However, not all segmentation techniques are suitable for disparity compensated stereo coding due to the excessive number of bits required to describe the shape and location of each region. A large amount of overhead is unacceptable in stereo coding applications where the disparity

number of picture coding applications. The traditional construction of the quad-tree representation starts with the assumption that the whole image can be represented by only one node (root) and an initial hypothesis test decides if further splitting is necessary. However, experimental results for both still and moving pictures have shown that in practise it is preferable to start by testing smaller blocks (typically 32x32 pixels) instead of the entire picture since the homogeneity test within larger blocks is rarely successful.

Similar constraints are introduced for the size of the smallest blocks in order to maintain the overhead information within acceptable limits. Research on variable transform coding and vector quantization suggests that the lowest level of the quad-tree representation should be in the region of 4x4 pixels. In the case of disparity compensation however, this size is too small and requires an unacceptably large amount of addressing information (overhead). In an attempt to avoid overshooting of the overhead information, a minimum permitted block size of 8x8 pixels is preferred. To further simplify the segmentation process, the AC energy of each pixel block may be used as the hypothesis test. The AC energy is defined as:

$$AC.energy=\Sigma_{i=1}^{N^2}(g_i-\bar{g})^2$$

where g_i is the intensity values of the individual pixels in the block and $\bar{g}$ is the mean intensity value of the whole block. According to this approach, the algorithm calculates the AC energy for each NxN pixel block and if its value is greater than a threshold, further subdivision of the block is carried out. The threshold value is suitably chosen to be

predetermined pattern (i.e. 9 search positions as shown in Fig.4). As noted earlier, there are 6561 candidate quadrilaterals for each block. Assuming that 3 bits are required to describe the displacement of each corner, there are 12 bits per block to describe the mapping parameters to the decoder. This relatively high figure represents the upper limit for the disparity overhead. Adoption of a more sophisticated coding scheme than a simple PCM used here can further reduce this overhead.

An alternative way to reduce the overhead information is to impose a discrimination process which will compensate only blocks that exhibit considerable improvement after the application of the generalized block matching. Details of such implementation is described below.

The two-stage algorithm described above is applied on blocks of differing sizes which has the advantage of compensating more accurately those parts of the screen having larger disparity values due to the smaller size of the blocks allocated to those areas of the picture. On the other hand, large blocks are assigned to low disparity areas which are usually successfully compensated with only the translational component of the disparity vector. This is in accordance with the characteristics of the Human Visual System (HVS) regarding the depth resolution required for an accurate perception of 3D from stereopair images. Hence, the size of each block also gives cues to whether the application of the computationally expensive generalized block matching is necessary or not.

As an example of the application of the two-stage algorithm described above, Fig.7 shows the absolute difference between two members of a stereo pair of images of an engine component (for which the quad-

number of bits required to code the right eye picture using the method of the present invention in the example was 325993 bits in total. This compares very favourably with the 386467 bits required to code the same picture using a simpler prediction formed by the conventional full search block matching with fixed sized blocks (16x16 pixels) and utilizing the same search window of 16x9 pixels.

We have found that the generalized block matching using quad-tree decomposition results in lower bit-rate than conventional block matching methods because it produces a better prediction. To verify this statement, we have compared the signal to noise ratio (SNR) of the reconstructed predictors measured with reference to the right-eye image (see Table 1). The prediction from the generalized variable block matching is better overall than that of the conventional fixed size block matching especially in occluded areas and areas of high curvature. Moreover, the quad-tree segmentation directs the computational efforts to the most significant parts of the picture thereby providing considerable savings in overhead information.

As previously mentioned area-based stereo disparity estimation shares many characteristics with motion compensation applied to consecutive frames of an image sequence. Both methods segment an image into fixed sized rectangular blocks and assume that each block is undergoing independent uniform translation given by a displacement vector $V=(dx,dy)$. For each block in one image (i.e. previous frame in motion estimation or right-eye image in disparity estimation) a thorough comparison with all possible corresponding blocks is performed, within a search area in the other image. The best match is found by minimising a distortion measurement, or by maximising a correlation

other features which are already known in the design, manufacture and use of image transmission and storage systems, display apparatuses and component parts thereof and which may be used instead of or in addition to features already described herein. Although claims have been formulated in this application to particular combinations of features, it should be understood that the scope of the disclosure of the present invention also includes any novel feature or any novel combination of features disclosed herein either explicitly or implicitly or any generalisation thereof, whether or not it relates to the same invention as presently claimed in any claim and whether or not it mitigates any or all of the same technical problems as does the present invention. The applicants hereby give notice that new claims may be formulated to such features and/or combinations of features during the prosecution of the present application or of any further application derived therefrom.

CLAIMS

1. A method for disparity encoding of a first two-dimensional image in relation to a second two-dimensional image, in which each of the first and second images is divided into a plurality of regular non-overlapping blocks and, for each block of the first image;

a) comparing the block with those blocks at and near the corresponding position in the second image and selecting that providing the best match;

b) applying a predetermined series of distortions to the selected block of the second image, comparing the result of each distortion with the first image block, and identifying the applied distortion providing the best match to the first image block: and

c) storing the location of the selected second image block and applied distortion,

characterised in that the first and second images are initially divided into blocks of a first size, each block of the first image is compared with the correspondingly positioned block of the second image in accordance with predetermined matching criteria and, if the criteria are not met, the first and second image blocks are divided into corresponding sub-blocks and each sub-block compared according to the same criteria, the step of dividing into sub-blocks and comparing being repeated until the criteria are satisfied, and steps a),b) and c) are then performed for each block or sub-block meeting the criteria.

2. A method according to Claim 1, in which the step of dividing a block into sub-blocks, or a sub-block into further sub-blocks, comprises dividing the block or sub-block into four equal portions.

8. Disparity encoding apparatus operable to receive first and second two-dimensional images and to encode the first image terms of its disparity with respect to the second image, the apparatus comprising: image receiving means arranged to receive the first and second images and to divide each into a plurality of corresponding non-overlapping blocks of a first size;
first comparison means operable to compare each block of the first image with the corresponding block of the second image and a plurality of blocks surrounding the said corresponding block in accordance with predetermined matching criteria and, where the criteria are not met, operable to indicate so to the image receiving means, the image receiving means thereafter operating to divide the block failing to meet the predetermined matching criteria together with the corresponding block of the second image into a plurality of sub-blocks, with the first comparison means being arranged to then repeat the comparison for each sub-block in accordance with the predetermined matching criteria;
image modulation means operable to apply, for each first image block or sub-block, a predetermined series of distortions to the respective selected second image;
second comparison means arranged to compare each of the series of distorted versions of the selected second image block or sub-block to the respective first image block or sub-block, in accordance with a second predetermined matching criteria;and
storage means connected to the first and second comparison means and operable to receive and store, for each block or sub-block of the first image, information identifying the block of the second image meeting the predetermined matching criteria, and the applied distortion best satisfying the second predetermined matching criteria.

26

Patents Act 1977
E: niner's report to the Comptroller under Section 17
(Th. Search report)

Application number
GB 9326582.5

Relevant Technical Fields

(i) UK Cl (Ed.N) H4F (FDD, FGM, FRG, FRP, FRW, FRX)

(ii) Int Cl (Ed.5) H04N (5/14, 7/137, 13/00, 15/00) G06F (15/70)

Search Examiner
M K REES

Date of completion of Search
26 JANUARY 1995

Databases (see below)
(i) UK Patent Office collections of GB, EP, WO and US patent specifications.

(ii) ONLINE DATABASE: WPI

Documents considered relevant following a search in respect of Claims :-
1 to 10

Categories of documents

X: Document indicating lack of novelty or of inventive step.

Y: Document indicating lack of inventive step if combined with one or more other documents of the same category.

A: Document indicating technological background and/or state of the art.

P: Document published on or after the declared priority date but before the filing date of the present application.

E: Patent document published on or after, but with priority date earlier than, the filing date of the present application.

&: Member of the same patent family; corresponding document.

Category	Identity of document and relevant passages		Relevant to claim(s)
A	GB 2198310 A	(BBC) see abstract	1, 8
A	EP 0353644 A2	(SCHLUMBERGER TECHNOLOGIES) see whole document	1, 8
A	EP 0254643 A2	(FAIRCHILD SEMICONDUCTOR CORP) see abstract	1, 8
A	EP 0005918 A1	(HUGHES AIRCRAFT COMPANY) see abstract	1, 8

Databases:The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal (Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).